

ISSN (print): 2421-5783
ISSN (on line): 2421-5562



Consiglio Nazionale delle Ricerche

IRCES

ISTITUTO DI RICERCA SULLA CRESCITA ECONOMICA SOSTENIBILE
RESEARCH INSTITUTE ON SUSTAINABLE ECONOMIC GROWTH

Rapporto Tecnico

Numero 4, Luglio 2017

Archivio Studi Adriatici (ASA) al servizio
della ricerca: istruzioni per l'uso

***Simona Armeli Minicante, Giancarlo Birello,
Alessandro Ceregato, Anna Perin***



Direttore Secondo Rolfo

Direzione CNR-IRCRES
Istituto di Ricerca sulla crescita economica sostenibile
Via Real Collegio 30, 10024 Moncalieri (Torino), Italy
Tel. +39 011 6824911 / Fax +39 011 6824966
segreteria@ircres.cnr.it
www.ircres.cnr.it

Sede di Roma Via dei Taurini 19, 00185 Roma, Italy
Tel. +39 06 49937809 / Fax +39 06 49937808

Sede di Milano Via Bassini 15, 20121 Milano, Italy
Tel. +39 02 23699501 / Fax +39 02 23699530

Sede di Genova Università di Genova Via Balbi, 6 - 16126 Genova
Tel. +39 010 2465459 / Fax +39 010 2099826

Redazione Secondo Rolfo (direttore responsabile)
Antonella Emina
Anna Perin
Enrico Viarisio
Isabella Maria Zoppi
 redazione@ircres.cnr.it
 www.ircres.cnr.it/index.php/it/produzione-scientifica/pubblicazioni

RAPPORTO TECNICO CNR-IRCRES, numero 4, Luglio 2017



Copyright © Luglio 2017 by CNR - IRCRES

Archivio Studi Adriatici (ASA) al servizio della ricerca: istruzioni per l'uso*

Archivio Studi Adriatici (ASA) at the research service: user's instructions

SIMONA ARMELI MINICANTE^a, GIANCARLO BIRELLO^b, ALESSANDRO CEREGATO^a, ANNA PERIN^b

^a CNR-ISMAR, Consiglio Nazionale delle Ricerche, Istituto di Scienze Marine, Arsenale - Tesa 104, Castello 2737/F, 30122 Venezia – Italy

^b CNR-IRCRES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, via Real Collegio 30, Moncalieri (TO) – Italy

corresponding author: simona.armeli@ve.ismar.cnr.it

ABSTRACT

The *Archivio di Studi Adriatici* (ASA) is a repository of the Institute of Marine Sciences (ISMAR-CNR) of Venice, completely open source and open access. It was developed in cooperation with the IRCRES-CNR of Turin with the aims to preserve and to make accessible the natural collections, heritage books, documents and maps of the Institute of Marine Sciences. Digitized specimens and metadata, compiled using Dublin Core and Simple Darwin Core formats, are preserved into a Fedora Repository, public accessed by Islandora framework and reachable through the website www.archiviostudiadriatici.it.

KEYWORDS

Repository, Open Access, digital preservation, natural collection

JEL CODES: Y91, Y92

DOI: 10.23760/2421-5562.2017.004

HOW TO CITE THIS ARTICLE

Armeli Minicante Simona, Birello Giancarlo, Ceregato Alessandro, Perin Anna, 2017. "Archivio Studi Adriatici (ASA) al servizio della ricerca: istruzioni per l'uso", *Rapporto Tecnico IRCrES-CNR*, n. 4/2017.

*Si ringraziano il dott. Fabio Trincardi e il dott. Secondo Rolfo per aver sostenuto la collaborazione tra i due istituti CNR per la realizzazione di questo progetto.

INDICE

1	PERCHÉ ASA.....	3
1.1	ASA il portale	4
1.2	ASA il repository	4
1.3	Le policy	4
2	LE COLLEZIONI PRESENTI	7
3	COME SONO ORGANIZZATE LE COLLEZIONI	8
4	L'IMPORTANZA DEI METADATI.....	8
4.1	Metadati Dublin Core	9
4.2	Metadati Darwin Core	12
5	AGGIUNGERE UNA COLLEZIONE.....	16
5.1	Catalogazione.....	16
5.2	Digitalizzare, regole e suggerimenti	16
5.3	Metadatazione, organizzazione del foglio elettronico	16
5.4	Trasmissione delle scansioni e dei metadati per la messa in rete	18
6	BIBLIOGRAFIA	20
7	GLOSSARIO.....	21

Archivio Studi Adriatici (ASA) al servizio della ricerca: istruzioni per l'uso

SIMONA ARMELI MINICANTE, GIANCARLO BIRELLO, ALESSANDRO CEREGATO,
ANNA PERIN

1 PERCHÉ ASA

L'Archivio di Studi Adriatici, conservato presso la sede storica del complesso Canonica di Venezia, raccoglie il patrimonio storico degli Istituti di Ricerca che qui hanno avuto sede a partire dal 1932: il Regio Comitato Talassografico Italiano e l'Istituto di Studi Adriatici, l'Osservatorio della Pesca Marittima, l'Istituto di Biologia del Mare e infine l'Istituto per lo Studio della Dinamica delle Grandi Masse, fino alla fusione, nel 2003, nell'Istituto di Scienze Marine (ISMAR-CNR) che oggi ha la sua sede principale presso l'Arsenale di Venezia.

Nel corso degli anni, i numerosi progetti di ricerca condotti da ISMAR hanno portato alla costituzione di raccolte di organismi provenienti dai diversi ambienti estuarini e marini del Mediterraneo. Ad esse si affiancano le ricerche geologiche iniziate nei primi anni '70 per indagare le cause della subsidenza della laguna di Venezia, le cui carote estratte dai sondaggi costituiscono uno straordinario archivio dell'evoluzione geologica ed ambientale di quest'area nel Pleistocene.

La sezione più ricca e probabilmente la più preziosa riguarda tuttavia i fondi accumulati durante l'attività degli istituti legati al Comitato Talassografico Italiano che avevano sede presso la residenza che lo scultore Pietro Canonica decise di donare a Guglielmo Marconi nel 1932 con il preciso vincolo di farne un istituto per lo studio del mare sotto ogni profilo. A seguito dell'insediamento dell'Istituto di Studi Adriatici (1933) fu creata una biblioteca multidisciplinare dedicata alla conoscenza dell'Adriatico, della laguna di Venezia e dei Balcani, di cui l'Italia rivendicava l'annessione. Nel 1939 all'ISA fu affiancato l'Osservatorio di Pesca Marittima che nel 1942-43 raccolse anche i ricercatori della sede omologa di Rovigno, mentre dal 1946 in poi l'ex Comitato Talassografico (poi Centro Nazionale di Studi Talassografici) raccolse progressivamente l'eredità degli istituti precedenti fino alla riorganizzazione dell'Istituto di Biologia del Mare nel 1968. Infine, il Laboratorio (poi Istituto) per lo studio della Dinamica delle Grandi Masse, istituito nel 1969 a Palazzo Papadopoli in Canal Grande, trasferì la propria biblioteca presso la sede dell'Istituto di Biologia del Mare dopo la fusione in ISMAR del 2002-2003.

A testimonianza di questa storia complessa e malgrado l'eliminazione e la dispersione di molti materiali nel corso degli anni, i fondi librari storici sono oggi costituiti da oltre 10000 titoli fra libri (i più antichi risalgono al Seicento) e riviste e da una collezione di circa 5000 tra mappe e portolani che include un ampio repertorio di esemplari antichi risalenti anche alla seconda metà del XVI secolo. La storia e lo sviluppo di questi istituti sono contenuti in un ricchissimo archivio di manoscritti, corrispondenze, fotografie e documenti a partire dall'istituzione del Comitato Talassografico Italiano. Tra le testimonianze di storia materiale degli istituti, oltre ad una serie di strumenti sopravvissuti a traslochi e dismissioni, un erbario algologico raccolto dall'Osservatorio della Pesca e dall'Istituto di Studi Adriatici tra il 1941 e il 1950, è giunto a noi in ottime condizioni.

I progetti per la valorizzazione delle collezioni stanno diventando sempre più frequenti negli istituti di ricerca e il loro uso e divulgazione è notevolmente facilitato dall'introduzione e dallo sviluppo della tecnologia informatica e delle banche dati.

Le raccolte di ISMAR, eterogenee per origine, epoca e disciplina, sono accomunate da un tema comune: lo studio del mare e dei casi particolari della Laguna di Venezia e dell'Adriatico, per cui rappresentano un repertorio forse unico in Italia per specificità e ricchezza e difficilmente confrontabile con archivi analoghi in giro per il Mondo. Un patrimonio scientifico da conservare, valorizzare e diffondere, anche alla luce del recente accordo di collaborazione fra MiBact, MIUR, ANCI e ANMS, volto alla tutela, promozione e diffusione del patrimonio culturale e tecnico-scientifico italiano.

Grazie alla collaborazione con l'Istituto di Ricerca sulla Crescita Economica Sostenibile IRCrES CNR di Moncalieri (TO), l'Archivio di Studi Adriatici (ASA) costituisce il portale digitale open source di ISMAR Venezia per la consultazione di libri, carte, documenti e collezioni scientifiche della sede storica dell'Istituto.

1.1 ASA il portale

All'indirizzo www.archiviostudiadriatici.it è stato realizzato un sito web responsive in Joomla utilizzando il template free *jpeople* di *Freshjoomla templates*. Scopo del sito è fungere da portale di presentazione del progetto con l'indicazione dei contatti e l'indirizzamento verso il repository. In particolare sono stati creati dei bottoni specifici per orientare e veicolare gli utenti verso le diverse collezioni dando contemporaneamente una visione di insieme per stimolare la consultazione dei diversi archivi.

Dal portale, inoltre, è possibile accedere alle pagine delle iniziative a cui aderiscono le collezioni dell'Archivio di Studi Adriatici (es. Global Registry of Scientific Collections, CollMap e Index Herbariorum).

1.2 ASA il repository

L'architettura software di conservazione è basata su Fedora Repository, ampiamente usato per la gestione di oggetti digitali e annovera tra le sue caratteristiche il supporto di grandi quantità di dati, l'assegnazione di identificatori persistenti, API per l'ingesting programmabile, la descrizione semantica delle relazioni tra gli oggetti e la loro gestione basata su modelli. A questo componente è stato affiancato SOLR, un altro software open source, che fornisce una piattaforma di indicizzazione e ricerca pronto all'uso con ottime prestazioni e alta efficienza. Completano il back-end BlazeGraph per la gestione delle triple e Adore-Djatoka quale image server per la gestione delle immagini in formato JPEG2000.

Lato utente, il software di presentazione è costituito da due principali applicazioni installate su un server Apache: il CMS Drupal che ospita al suo interno il framework open-source Islandora. Islandora mette a disposizione pacchetti specifici che consentono di gestire diversi tipi di dati relativi a settori scientifici differenti (ad esempio mappe, immagini, libri). Islandora include una serie di strumenti che dialogano direttamente con Fedora Repository, permettendo di accedere agli oggetti e ai loro metadati e consentendo di renderli visibili all'utente in forma organizzata.

L'architettura hardware si sviluppa su una piattaforma virtualizzata che ospita la macchina virtuale del repository e quella del front-end con relativa interfaccia utente di accesso ai dati. A complemento delle macchine virtuali uno storage (SAN) ad alta affidabilità per la memorizzazione degli oggetti digitalizzati ed il backup dei componenti software.

1.3 Le policy

Per la redazione delle policy pubblicate sul repository è stato utilizzato il tutorial messo a disposizione dal sito di OpenDOAR (The Directory of Open Access Repositories).

Il tutorial permette, con particolare attenzione agli standard Open Access e, semplicemente utilizzando le caselle di controllo e gli elenchi di selezione, di formulare e presentare le policy sul riutilizzo degli elementi depositati, sia metadati che dati, definendo inoltre le policy del 'contenitore' repository.

Nelle policy viene anche dichiarato che ASA ha effettuato ogni ragionevole sforzo per rintracciare gli eventuali titolari di diritti d'autore e ottenere l'autorizzazione alla riproduzione delle opere a scopo non commerciale. Non sempre ciò è stato possibile. ASA è tuttavia pronta a rimuovere in ogni momento le opere digitalizzate dietro richiesta documentata da parte dei titolari di eventuali diritti.

Le policy del repository sono in lingua inglese, disponibili on line a questa pagina <http://asa.archiviostudiadriatici.it/policy> e qui riportate:

Metadata Policy

for information describing items in the repository

1. *Anyone may access the metadata free of charge.*
2. *The metadata may be re-used in any medium without prior permission for not-for-profit purposes provided the OAI Identifier or a link to the original metadata record are given.*
3. *The metadata must not be re-used in any medium for commercial purposes without formal permission.*

Data Policy

for full-text and other full data items

1. *Anyone may access full items free of charge.*
2. *Single copies of full items can be:*
 - *reproduced, displayed or performed, and given to third parties in any format or medium*
 - *for personal research or study, educational, or not-for-profit purposes without prior permission or charge.*
3. *provided:*
 - *the authors, title and full bibliographic details are given*
 - *a hyperlink and/or URL are given for the original metadata page*
 - *the content is not changed in any way*
4. *Full items must not be sold commercially in any format or medium without formal permission of the copyright holders.*
5. *This repository is not the publisher; it is merely the online archive.*
6. *Mention of Archivio Studi Adriatici is appreciated but not mandatory.*

Content Policy

for types of document & data set held

1. *This is a multi-institution subject-based repository.*
2. *Subject Specialities:*
 - *Science General*
 - *Biology and Biochemistry*
 - *Geography and Regional Studies*
 - *History and Archaeology*
 - *Earth and Planetary Sciences*
 - *Language and Literature*
3. *Archivio Studi Adriatici is restricted to:*
 - *Books, chapters and sections*
 - *Multimedia and audio-visual materials*
 - *Other special item types*
 - *Special materials: maps, natural collection*
4. *Items are individually tagged with their version type and date.*
5. *Principal Languages: Italian*

Submission Policy

concerning depositors, quality & copyright

1. *Items may only be deposited by accredited members of the organisation, or their delegated agents.*
2. *The administrator only vets items for valid layout & format*
3. *The validity and authenticity of the content of submissions is the sole responsibility of the depositor.*
4. *No embargo policy defined.*
5. *Any copyright violations are entirely the responsibility of the authors/depositors.*
6. *If Archivio Studi Adriatici receives proof of copyright violation, the relevant item will be removed immediately.*

Preservation Policy

1. *Items will be retained indefinitely.*
2. *Archivio Studi Adriatici will try to ensure continued readability and accessibility.*
 - *Items will be migrated to new file formats where necessary.*
 - *Where possible, software emulations will be provided to access un-migrated formats.*
3. *Archivio Studi Adriatici regularly backs up its files according to current best practice.*
4. *Items may be removed at the request of the author/copyright holder, but this is strongly discouraged.*
5. *Acceptable reasons for withdrawal include:*
 - *Proven copyright violation or plagiarism*
 - *Legal requirements and proven violations*
 - *National Security*
 - *Falsified research*
6. *Withdrawn items are deleted entirely from the database.*
7. *Withdrawn items' identifiers/URLs are retained indefinitely.*
8. *The metadata of withdrawn items will not be searchable.*
9. *If necessary, an updated version may be deposited.*
 - *The item's persistent URL will always link to the latest version.*
10. *In the event of Archivio Studi Adriatici being closed down, the database will be transferred to another appropriate archive.*

Inoltre tutte le immagini relative alle collezioni naturalistiche (es. l'algaro) sono rilasciate con licenza Creative Common CC BY. Ossia l'utente è libero di:

- *Condividere* — riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato
- *Modificare* — remixare, trasformare il materiale e basarti su di esso per le tue opere
- *per qualsiasi fine, anche commerciale.*

Il licenziante non può revocare questi diritti fintanto che vengano rispettati i termini della licenza alle seguenti condizioni:

- *Attribuzione* — Devi riconoscere una menzione di paternità adeguata, fornire un link alla licenza e indicare se sono state effettuate delle modifiche. Puoi fare ciò in qualsiasi maniera ragionevole possibile, ma non con modalità tali da suggerire che il licenziante avalli te o il tuo utilizzo del materiale.

- Divieto di restrizioni aggiuntive — Non puoi applicare termini legali o misure tecnologiche che impongano ad altri soggetti dei vincoli giuridici su quanto la licenza consente loro di fare.

2 LE COLLEZIONI PRESENTI

Le collezioni presenti nel repository ASA comprendono immagini di campioni biologici, libri e mappe.

a. Algarium Veneticum

L'*Algarium Veneticum* (IH: ISMAR) è l'erbario istituzionale dedicato alla vegetazione marina. Attualmente comprende:

- la collezione Minio e Spada: realizzata da Michelangelo Minio (1872-1960), botanico e naturalista, con il contributo di Nicolò Spada e Giacomo Zolezzi, e comprendente in totale 1169 fogli d'erbario (*exsiccata*). La sezione principale è intitolata "*Distribuzione e polimorfismo di Gracilaria confervoides nella laguna di Venezia*" e contiene 19 carpette con 884 *exsiccata* identificati dagli Autori come *Gracilaria confervoides* (L.) Greville. I campioni sono stati raccolti tra il 1941 e il 1950 in 107 stazioni di campionamento distribuite nel centro storico della città di Venezia, nelle isole di Chioggia, Lido e Murano. La collezione comprende anche una sezione di miscellanea costituita da 9 carpette contenenti 285 *exsiccata* appartenenti a differenti taxa di alghe rosse, verdi e brune, raccolte nello stesso periodo nella Laguna di Venezia;
- un esemplare di *Corallina mediterranea*: raccolta nel 1870 a Biarritz (Francia) e proveniente dallo storico Herbarium Thuret, rappresenta il campione più antico dell'*Algarium Veneticum*;
- la collezione di macroalghe della laguna di Venezia: costituita da 197 *exsiccata* di Rhodophyta (alghe rosse), 34 di Phaeophyceae (alghe brune) e 164 di Chlorophyta (alghe verdi) raccolte tra il 2014 e il 2016.

Tutti i campioni storici e parte di quelli moderni sono già disponibili sul repository ASA.

b. Biblioteca Studi Adriatici

Il fondo storico è costituito dalla biblioteca dell'Istituto di Studi Adriatici, creata nel 1933 come primo atto dell'Istituto e dedicata in massima parte allo studio di Venezia e della sponda orientale dell'Adriatico, con particolare riferimento alla storia, all'etnografia e alle lingue balcaniche. Comprende circa 2000 volumi risalenti anche al Seicento tra cui una Dissertazione delle Antichità d'Italia di Lodovico Antonio Muratori nell'edizione del 1765-66, una serie quasi completa della Gazzetta di Venezia dal 1816 al 1945, una raccolta di Leggi e Decreti dall'ingresso delle Truppe Austriache a Venezia nel 1798 all'Unità d'Italia e numerose monografie dedicate a Venezia, ai Balcani, alla navigazione ed alle attività marittime in laguna e in Adriatico. Il repository attualmente contiene 747 titoli dal fondo ex ISA.

c. Cartoteca storica

L'Istituto di Studi Adriatici fin dai primi anni di attività dedicò una particolare attenzione alla raccolta di immagini, mappe e vedute prospettiche di Venezia e del territorio della Repubblica, dei Balcani e dell'Adriatico, ma non si limitò a questi soli argomenti. Il Registro Topografico riporta 107 carte che vanno dall'Italia Meridionale di Pirro Ligorio e Ortelius della seconda metà del XVI secolo alle mappe dell'Albania redatte negli anni '40 del secolo scorso, un portolano olandese del 1654, un portolano veneto del 1790-'92 e un esemplare quasi perfetto dell'Atlante Veneto di Vincenzo Coronelli, pubblicato nel 1691. Oltre a questo materiale storico, che si compone di esemplari spesso unici come la *Corografia della Dalmatia* (Ceregato, 2015), tracciata da Giusto Emilio Alberghetti ad inchiostro e acquerello sulla base dei rilievi di Luigi Ferdinando Marsili, che riporta gli esiti dei trattati di Carlowitz del 1699, e di una grande

planimetria di Corfù della seconda metà del Seicento, sempre manoscritta, la cartoteca raccoglie una ricca collezione di carte topografiche, geologiche, nautiche e catastali dal 1800 al 2002, appartenute ai diversi istituti che si sono susseguiti con sede a Venezia fino a ISMAR.

In seguito ad una campagna di catalogazione, digitalizzazione e georeferenziazione delle mappe storiche condotta in collaborazione con il Laboratorio Cartografico CIRCE - IUAV di Venezia, è stato creato il portale cartografico BSA <http://cigno.ve.ismar.cnr.it/bsa/>, incluso nell'Atlante della Laguna <http://www.atlantedellalaguna.it>, attraverso cui è possibile consultare una prima parte delle mappe georeferenziate o geolocalizzate in formato vettoriale e confrontarle fra loro e con mappe più recenti (Balletti et al., 2016). È allo studio la connessione tra le mappe in formato vettoriale e le corrispondenti digitalizzazioni in formato originale sul repository ASA.

3 COME SONO ORGANIZZATE LE COLLEZIONI

Nell'ambito del repository, per 'collezione' si intende un insieme organico e congruente di oggetti, come ad esempio una raccolta di volumi sulla tecnica del '700, un insieme di immagini di un particolare organismo o la raccolta di mappe della Laguna di Venezia.

L'insieme collezione ha la peculiarità di poter essere trattato in blocco e costituisce l'oggetto indivisibile per alcune operazioni quali: ordinamento, set per l'OAI-PMH, visualizzazioni specifiche, insieme per la metadattazione.

Una collezione può contenere oggetti e/o ulteriori collezioni, ma è buona norma, per un'agevole navigazione all'interno del repository da parte dell'utente, non annidare troppi livelli. Per questa ragione si è deciso inizialmente di avere al massimo due livelli di gerarchia: il repository conterrà una serie di collezioni principali che potranno avere a loro volta una sola sotto-collezione.

Quando si naviga nel repository, le collezioni vengono visualizzate come insiemi paginati di thumbnails e relativi titoli, il default di visualizzazione è 20 oggetti per pagina. L'ordinamento tipico disponibile è per titolo o per data, crescente o decrescente, o una combinazione di questi, mentre per casi più specifici andrà valutato di volta in volta. Sul margine destro può essere attivato, per collezione, un elenco linkato estensibile di tutti gli oggetti della collezione, raggruppati ad esempio per anno o per decade, fornendo un pratico punto di accesso ai singoli oggetti della collezione.

Non sono invece dipendenti dalla collezione di appartenenza i metadati assegnati ai singoli oggetti, cioè significa che all'interno di una collezione di immagini possono coesistere immagini corredate dai soli Dublin Core con quelle corredate anche dei Darwin Core.

Per quanto riguarda ad esempio le collezioni dell'*Algarium Veneticum*, si è scelto di rendere immediatamente visibili alcuni Dublin Core e Darwin Core associati alle immagini, al fine di fornire le informazioni più esaustive sull'oggetto, senza appesantire la consultazione per l'utente. Sul margine sinistro è invece disponibile un elenco linkato dal quale è possibile fare il download dell'immagine (in formato TIFF e JPG), scaricare l'XML contenente tutti i metadati associati all'immagine (sia Dublin che Darwin Core), accedere a pagine web esterne inerenti (es. Algaebase).

4 L'IMPORTANZA DEI METADATI

La descrizione degli oggetti della collezione viene fornita dall'insieme di diverse tipologie di informazione indicate con il nome generico di *metadati* (dal greco μετά "oltre, per mezzo" e dal latino *datum* "informazione", letteralmente "dato per mezzo di un altro dato").

Al fine di creare una qualsiasi collezione digitale, è necessario verificare come il materiale sia originariamente organizzato e visualizzato (Schmidt, 2007). Infatti, in base al tipo di oggetto e quindi alla tipologia di informazione che esso porta con sé, è opportuno scegliere l'organizzazione più adatta di metadati. Per questo motivo, essendo l'Archivio di Studi Adriatici un repository contenente collezioni di varia natura, si è scelto di integrare allo standard dei Du-

blin Core quello dei Darwin Core per creare un set di metadati completo e rilevante sia per i biologi sia per i bibliotecari (Biodiversity Information Standards – TDWG, 2015).

4.1 Metadati Dublin Core

Il “Dublin Core Metadata Element Set” è un vocabolario di quindici elementi di dati opzionali (Caplan, 2003, vedi tabella) e utilizzabili per descrivere un’ampia gamma di oggetti. Per la loro versatilità (è anche possibile introdurre valori multipli per ogni elemento) sono utili in diversi contesti biblioteconomici ed archivistici.

Gli elementi Dublin Core descritti in questo standard fanno parte di un più ampio insieme di vocabolari di metadati e specifiche tecniche gestite dal Dublin Core Metadata Initiative (DCMI). Il set completo dei vocabolari (DCMI Metadata Terms) include anche set di classi di risorsa, schemi di codifica vocabolario e schemi di codifica di sintassi. Tutti questi possono essere utilizzati in combinazione con altri vocabolari compatibili.

Nel nostro caso i quindici elementi base, (Tabella 1), sono più che sufficienti a coprire le esigenze di metadatazione per quanto riguarda i libri.

Tabella 1. Elementi Dublin Core

Term Name: contributor	
URI:	http://purl.org/dc/elements/1.1/contributor
Label:	Contributor
Definition:	An entity responsible for making contributions to the resource.
Comment:	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Term Name: coverage	
URI:	http://purl.org/dc/elements/1.1/coverage
Label:	Coverage
Definition:	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Comment:	Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.
References:	[TGN] http://www.getty.edu/research/tools/vocabulary/tgn/index.html
Term Name: creator	
URI:	http://purl.org/dc/elements/1.1/creator
Label:	Creator
Definition:	An entity primarily responsible for making the resource.

Comment:	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Term Name: date	
URI:	http://purl.org/dc/elements/1.1/date
Label:	Date
Definition:	A point or period of time associated with an event in the lifecycle of the resource.
Comment:	Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
References:	[W3CDTF] http://www.w3.org/TR/NOTE-datetime
Term Name: description	
URI:	http://purl.org/dc/elements/1.1/description
Label:	Description
Definition:	An account of the resource.
Comment:	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
Term Name: format	
URI:	http://purl.org/dc/elements/1.1/format
Label:	Format
Definition:	The file format, physical medium, or dimensions of the resource.
Comment:	Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].
References:	[MIME] http://www.iana.org/assignments/media-types/
Term Name: identifier	
URI:	http://purl.org/dc/elements/1.1/identifier
Label:	Identifier
Definition:	An unambiguous reference to the resource within a given context.
Comment:	Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.
Term Name: language	
URI:	http://purl.org/dc/elements/1.1/language
Label:	Language
Definition:	A language of the resource.

Comment:	Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].
References:	[RFC4646] http://www.ietf.org/rfc/rfc4646.txt
Term Name: publisher	
URI:	http://purl.org/dc/elements/1.1/publisher
Label:	Publisher
Definition:	An entity responsible for making the resource available.
Comment:	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Term Name: relation	
URI:	http://purl.org/dc/elements/1.1/relation
Label:	Relation
Definition:	A related resource.
Comment:	Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Term Name: rights	
URI:	http://purl.org/dc/elements/1.1/rights
Label:	Rights
Definition:	Information about rights held in and over the resource.
Comment:	Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.
Term Name: source	
URI:	http://purl.org/dc/elements/1.1/source
Label:	Source
Definition:	A related resource from which the described resource is derived.
Comment:	The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Term Name: subject	
URI:	http://purl.org/dc/elements/1.1/subject
Label:	Subject
Definition:	The topic of the resource.
Comment:	Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.

Term Name: title	
URI:	http://purl.org/dc/elements/1.1/title
Label:	Title
Definition:	A name given to the resource.
Comment:	Typically, a Title will be a name by which the resource is formally known.
Term Name: type	
URI:	http://purl.org/dc/elements/1.1/type
Label:	Type
Definition:	The nature or genre of the resource.
Comment:	Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.
References:	[DCMITYPE] http://dublincore.org/documents/dcmi-type-vocabulary/

Fonte: (<http://dublincore.org/documents/dces/>)

4.2 Metadati Darwin Core

Nel caso di oggetti attinenti la biologia, i soli Dublin Core potrebbero non essere sufficienti per includere tutti i metadati opportuni e disponibili per la loro caratterizzazione e classificazione. L'architettura del repository è stata estesa per accettare anche l'eventuale aggiunta di un'altra classe di metadati, molto diffusi in campo biologico, i Darwin Core, che potranno essere attribuiti per oggetto, a prescindere dal suo formato multimediale.

I Darwin Core, basati sugli standard sviluppati dalla Dublin Core Metadata Initiative (DCMI, <http://dublincore.org/>), sono un insieme di termini aventi una semantica chiaramente definita in modo da essere compresi dalle persone o interpretati dalle macchine, rendendo in questo modo possibile determinare gli usi appropriati dei dati in esso codificati.

I termini sono organizzati in tredici classi, sei delle quali coprono aspetti generali (eventi, ubicazione, contesto geologico, occorrenza, taxon e identificazione) del dominio della biodiversità. Le categorie rimanenti coprono le relazioni con altre risorse, le misure e le informazioni generiche sui record.

Per rispondere ai requisiti del repository abbiamo implementato lo schema semplice dei Darwin Core (simple DWC), ossia una versione piatta dei termini Darwin Core.

Come esempio riportiamo i simple Darwin Core e gli elementi Dublin Core selezionati per le collezioni dell'*Algarium Veneticum* (Tabella 2).

Tabella 2. Elementi Darwin Core e Dublin Core utilizzati

Standard	Term	Definition
dc	title	A name given to the resource
dc	description	An account of the resource

dc	subject	The topic of the resource
dwc	institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record
dwc	recordNumber	An identifier given to the Occurrence at the time it was recorded. Often serves as a link between field notes and an Occurrence record, such as a specimen collector's number
dwc	collectionCode	The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived
dwc	order	The full scientific name of the order in which the taxon is classified
dwc	scientificName	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the identification-Qualifier term
dc	creator	An entity primarily responsible for making the resource
dwc	identifiedBy	A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the subject
dwc	acceptedNameUsage	The full name, with authorship and date information if known, of the currently valid (zoological) or accepted (botanical) taxon
dwc	taxonID	An identifier for the set of taxon information (data associated with the Taxon class). May be a global unique identifier or an identifier specific to the data set
dwc	eventDate	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E)
dc	date	A point or period of time associated with an event in the lifecycle of the resource

dwc	country	The name of the country or major administrative unit in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names
dwc	higherGeography	A list (concatenated and separated) of geographic names less specific than the information captured in the locality term
dwc	locationID	An identifier for the set of location information (data associated with dcterms:Location). May be a global unique identifier or an identifier specific to the data set
dwc	locality	The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description
dwc	decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive
dwc	decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive
dwc	coordinateUncertaintyInMeters	The horizontal distance (in meters) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term
dwc	georeferenceVerificationStatus	A categorical description of the extent to which the georeference has been verified to represent the best possible spatial description. Recommended best practice is to use a controlled vocabulary

dwc	geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude are based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown"
dwc	associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the Occurrence
dcterms	rightsHolder	A person or organization owning or managing rights over the resource
dc	publisher	An entity responsible for making the resource available
dc	rights	Information about rights held in and over the resource
dcterms	accessRights	Information about who can access the resource or an indication of its security status. Access Rights may include information regarding access or restrictions based on privacy, security, or other policies
dcterms	bibliographicCitation	A bibliographic reference for the resource as a statement indicating how this record should be cited (attributed) when used. Recommended practice is to include sufficient bibliographic detail to identify the resource as unambiguously as possible
dwc	associatedMedia	A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of media associated with the Occurrence
dc	type	The nature or genre of the resource
dcterms	type	The nature or genre of the resource. For Darwin Core, recommended best practice is to use the name of the class that defines the root of the record
dwc	basisOfRecord	The specific nature of the data record

Fonte: nostra elaborazione su tabelle elementi Dublin Core (<http://dublincore.org/documents/dces/>) e Darwin Core (<http://rs.tdwg.org/dwc/terms/>)

5 AGGIUNGERE UNA COLLEZIONE

Gli oggetti che andranno a popolare il repository dell'Archivio di Studi Adriatici seguiranno degli step ben definiti a prescindere del tipo di collezione trattata (campioni biologici, fogli d'erbario, libri, mappe, ecc.). In particolare per ogni oggetto sono previste le seguenti fasi:

Catalogazione

Digitalizzazione

Assegnazione metadati

Trasmissione del materiale per la messa in rete

5.1 Catalogazione

Ciascuna collezione viene catalogata secondo necessità interne alla collezione stessa. In linea generale ad ogni oggetto della collezione (*voucher*) viene associata un'etichetta cartacea che riporterà sinteticamente i principali dati del *voucher* (es., codice ID, nome scientifico della specie, data e luogo di raccolta, autori della collezione). Su un foglio elettronico sono contemporaneamente registrati, oltre ai dati riportati nell'etichetta cartacea, tutti gli altri dati disponibili (es., coordinate geografiche, status tassonomico, fenologia, ecc.). L'inserimento di queste informazioni permetterà di estrapolare con maggiore facilità i metadati necessari nella fase di ingestione.

5.2 Digitalizzare, regole e suggerimenti

Gli oggetti abitualmente da digitalizzare sono libri e immagini, in senso generico quindi per libro si intende qualsiasi tipo di pubblicazione cartacea (fascicolo, rivista, volume, collana, working paper, ...) e per immagine qualsiasi tipo di foto, disegno, mappa, diapositiva o simili. Un caso particolare rivestono invece le pubblicazioni già digitali che possono essere disponibili in vari formati, uno dei più comuni è sicuramente il pdf.

Il formato prescelto per la conservazione dei file, date le sue caratteristiche di conservazione di informazioni, è il TIFF.

Questo vale sia per i libri che per le immagini e anche per gli oggetti già digitali che andranno comunque convertiti dal formato originale a questo formato.

La risoluzione consigliata per la digitalizzazione delle immagini è di 400-600dpi, in funzione della dimensione, mentre per le pagine dei libri 300dpi è lo standard utilizzato, sufficiente per ottenere un'ottima risoluzione in lettura. Qualora le pagine contengano alcune illustrazioni particolarmente significative potrebbe essere opportuno effettuare una seconda scansione a parte a 400-600dpi e creare una collezione di immagini che potranno essere singolarmente collegate da relazioni alle relative pagine del libro.

È buona norma, nel caso dei libri in particolare, non rifilare troppo le scansioni, piuttosto è prudente lasciare i bordi (taglio anteriore, superiore e inferiore) rendendo più realistica la lettura on-line, mentre il taglio centrale delle pagine sarebbe opportuno non rifilarlo mai in modo che le due pagine affiancate nel viewer riproducano l'esatta percezione del libro cartaceo aperto.

Si ricorda che la digitalizzazione di un libro deve comprendere anche eventuali pagine bianche presenti in quanto la riproduzione deve rispecchiare esattamente la copia cartacea.

Il file prodotto dalla digitalizzazione dovrà quindi essere in formato TIFF, estensione .tif, non compresso e a colori (RGB).

5.3 Metadatazione, organizzazione del foglio elettronico

Gli oggetti digitali conservati nel repository sono molto vari e di diversa natura, per alcuni possono essere più che sufficienti i metadati Dublin Core (DC) mentre in altri casi è opportuno estendere le informazioni con altri schema come i Darwin Core (DWC).

Ad esempio immagini del settore biologico richiedono essere corredate dai DWC mentre per immagini tratte da un testo descrittivo o per delle mappe i DC potrebbero soddisfare le esigenze. Nel caso di libri e pubblicazioni generalmente i DC coprono tutte le necessità di metadatazione.

Nella trattazione seguente distingueremo tra 3 principali categorie di oggetti:

- BOOK: libri, riviste, qualsiasi tipo di pubblicazione assimilabile ad un libro e che sarà resa disponibile tramite un viewer on-line (Internet Archive Bookreader), normalmente disponibili i TIFF delle singole pagine;
- PDF: nel caso di oggetti già digitali, disponibili in PDF, saranno elaborati e trattati allo stesso modo dei BOOK;
- IMAGE: qualsiasi tipo di figura, illustrazione, immagine o assimilati, indipendente dai metadati richiesti, che sarà resa disponibile tramite un viewer on-line (OpenSeadragon), normalmente fornito il TIFF.

Per i nuovi oggetti, non ancora presenti nel repository, l'operazione di assegnazione dei metadati si sviluppa in due fasi. Inizialmente, prima dell'ingesting, gli oggetti dovranno essere corredati di un minimo di informazioni necessarie per un corretto inserimento nel repository, a tal scopo sarà fornito un foglio elettronico di base da riempire a cura del fornitore di contenuti. Una volta completato l'ingesting sarà reso disponibile un foglio elettronico nel quale inserire i metadati relativi agli oggetti precedentemente caricati nel repository. Il gestore del repository utilizzerà il foglio elettronico compilato dal fornitore di contenuti per popolare i metadati presenti nel repository.

Le informazioni iniziali richieste dipendono dal tipo di oggetto, in particolare sono:

- BOOK: nome cartella contenente i file, titolo, data, lingua, flag DWC (0/1);
- PDF: nome del file pdf senza estensione, titolo, data, lingua, flag DWC (0/1);
- IMAGE: nome del file, titolo, flag DWC (0/1).

Se, anche solo potenzialmente, gli oggetti dovranno essere corredati anche dai DWC andrà segnalato in anticipo valorizzando la relativa colonna del foglio elettronico di pre-ingesting.

Per i dettagli sui file e sull'organizzazione dei contenuti da trasmettere si rimanda al capitolo seguente.

I metadati sono scritti in file XML, poco 'umani', per questo si è deciso di utilizzare un foglio elettronico, di più facile compilazione, lasciando a degli script automatici la generazione dei file XML. Restano comunque alcune regole rigide da rispettare nella compilazione del foglio elettronico qui riassunte:

- una riga per oggetto;
- nelle celle NON inserire a capo, carattere paragrafo §, apice singolo ', e commerciale &, spazi in testa o in coda;
- il contenuto del foglio elettronico verrà convertito in XML, cioè un file di testo piatto, pertanto ogni formattazione (es. grassetto, colore, ...) è ininfluyente e verrà rimossa in fase di conversione;
- se un elemento non interessa o ha valore nullo, va lasciata vuota la relativa casella, ma NON va rimossa la colonna dell'elemento;
- se un elemento ha più valori, devono essere inseriti nella stessa cella separati da una barra verticale | senza spazi prima e dopo la barra (es. economia|automobile|filiera);
- il campo DATA (dc:date) va rigorosamente compilato rispettando la ISO 8601, cioè:
Anno:
AAAA (es. 1997)
Anno e mese:
AAAA-MM (es. 1997-07)
Anno, mese e giorno:
AAAA-MM-GG (es. 1997-07-16);
- la prima colonna PID e la colonna dc:identifier sono già compilate automaticamente e NON vanno mai modificate;
- la prima riga, relativa ai nomi dei campi, NON va mai modificata;

- la colonna lingua (dc:language) va valorizzata usando i codici ISO 639-2, tre lettere minuscole (ita, lat, eng, fra, deu,...);
- Se è presente più di un autore (dc:creator), vanno indicati come campi multipli separati da | (es. Mario Rossi|Giuseppe Verdi) utilizzando la forma Nome Cognome;
- anche le keywords (dc:subject) devono essere separate da una barra verticale | (es. operazioni navali|prima guerra mondiale);
- volendo assegnare un indice, nel caso di volumi, andrà inserito nel campo dc:description, in coda agli altri eventuali valori del campo, usando per ogni voce dell'indice il formato:

- Testo della voce dell'indice #6

- ✓ il trattino iniziale è il segno meno - e non il trattino lungo che a volte excel auto-sostituisce
- ✓ tra il trattino iniziale e il testo è obbligatorio uno spazio
- ✓ tra il testo ed il cancelletto è obbligatorio uno spazio
- ✓ tra il cancelletto ed il numero NON ci devono essere spazi
- ✓ Il testo va scritto tutto di seguito SENZA inserire ritorni a capo
- ✓ L'ordine con cui verranno presentate le voci corrisponde all'ordine con cui sono inserite nel campo description

Un esempio di dc:description completo di indice:

Idee, ricerche, programmi dell'Istituto Gramsci piemontese. Dicembre 1985|- Questioni di equità fiscale, intervista a Vincenzo Visco, a cura di Walter Santagata #1|- La forma di governo e i suoi paradossi, di Gustavo Zagrebelsky #6;

- nel campo dc:publisher va indicato nel caso dei volumi l'editore;
- nel campo dc:rights è possibile indicare il detentore del copyright;
- nel campo dc:type va indicata la tipologia dell'oggetto rispettando la terminologia del "DCMI Type Vocabulary" (es. Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text).

5.4 Trasmissione delle scansioni e dei metadati per la messa in rete

In generale si tratterà di inviare al gestore del repository l'insieme dei file relativi agli oggetti da inserire. La trasmissione potrà avvenire tramite invio fisico dell'HD contenente i file oppure a mezzo rete. In questo secondo caso il fornitore di contenuti potrà mettere a disposizione i file su un suo server per il download oppure caricarli su una partizione o un cloud messo a disposizione allo scopo, in ogni caso si valuterà caso per caso.

I file degli oggetti andranno forniti secondo uno schema predefinito in funzione della loro tipologia, nello specifico:

- NOTE Generali:
 - ✓ i nomi dei file e delle cartelle DEVONO essere SENZA spazi, SOLO numeri, lettere, () sottolineato, (-) segno meno o (.) punto
 - ✓ le estensioni dei file (.pdf, .txt, .tif) sempre 3 lettere in MINUSCOLO
 - ✓ il TITOLO fornito in questa fase potrà poi essere modificato liberamente nella seconda fase della metadattazione
 - ✓ la DATA, fornita in questa fase, è opportuno non modificarla più nel caso di BOOK e PDF in quanto verrà attribuita anche alle singole pagine e ne potrebbe falsare la ricerca, mentre nella seconda fase di metadattazione il campo DATA potrebbe invece essere meglio specificato aggiungendo anche mese e giorno se non inserito inizialmente, in ogni caso il valore della DATA dev'essere inserito rispettando la norma descritta al paragrafo precedente

- ✓ la LINGUA indicata in questa fase serve solo per l'OCR e l'HOCR e non interagisce in alcun modo coi metadati
- ✓ nel caso i file immagine a disposizione siano JPEG, in qualsiasi dei tre casi, andrà verificato previamente il grado di qualità e definito uno schema di consegna specifico
- ✓ in ogni caso è consigliabile un test preventivo sulla qualità e compatibilità dei file prima di procedere con la trasmissione delle intere collezioni
- BOOK
 - 1 cartella per collezione, contenente:
 - ✓ N cartelle, una per volume, ognuna contenente:
 - 1 file PDF comprensivo di tutte le pagine, possibilmente formato PDF/A, nome_qualsiasi.pdf
 - N file TIFF corrispondenti alle N pagine, formato TIFF non compresso a colori, nominati a partire da 0001.tif, 0002.tif, ...
 - (facoltativo) 1 cartella contenente eventuale OCR delle pagine, un file per pagina, formato txt UTF-8, nominati a partire da 0001.txt, 0002.txt, ...
 - ✓ foglio elettronico della collezione, una riga per volume, con le seguenti colonne
 - NOME CARTELLA
 - TITOLO
 - DATA
 - LINGUA
- PDF
 - Se fornito link ai file:
 - ✓ Foglio elettronico, un foglio x collezione, una riga per volume, con le seguenti colonne
 - LINK (es. http://mio.dominio.it/download/file_1.pdf)
 - TITOLO
 - DATA
 - LINGUA
 - se forniti file:
 - ✓ 1 cartella x collezione, contenente
 - i file, nome_qualsiasi.pdf
 - Un foglio elettronico per collezione, una riga per volume, con le seguenti colonne
 - NOME FILE
 - TITOLO
 - DATA
 - LINGUA
- IMAGE
 - ✓ 1 cartella per collezione, contenente:
 - i file delle immagini, nome_qualsiasi.tif
 - un foglio elettronico per collezione, una riga per immagine, con le seguenti colonne
 - NOME FILE
 - TITOLO

Dopo l'ingesting degli oggetti, verrà inviato al fornitore dei contenuti un foglio elettronico precompilato nel quale inserire i metadati. Il foglio conterrà già:

- il PID dell'oggetto, univoco all'interno del repository e costituito da una prima parte (namespace) separata dai due punti (:) da una seconda parte generalmente numerica;
- il titolo ricavato dalle informazioni pre-ingesting che potrà essere modificato;
- la data, nel caso di BOOK e PDF, da non modificare possibilmente almeno per quel che riguarda l'anno.

Nella compilazione, se le informazioni già presenti non sono sufficienti, si potrà consultare l'oggetto on-line inserendo il PID all'interno dell'URL, ad esempio se il PID è `libria:136195` nella barra di navigazione del browser va inserito:

`http://asa.archiviostudiadriatici.it/islandora/object/libria:136195`

Il foglio elettronico completato con i metadati dovrà essere restituito al gestore del repository che provvederà ad aggiornare i metadati degli oggetti in questione.

6 BIBLIOGRAFIA

- Abbà Simona, Birello Giancarlo, Vallino Marta, Perin Anna, Ghignone Stefano, Caciagli Piero, 2015. "Shall we share? A repository for Open Research Data in agriculture and environmental sciences", *Bulletin OEPP/EPPO Bulletin* vol. 45, n. 2, pp. 311-316.
- Armeli Minicante Simona, Sigovini Marco, Manghisi Antonio, Le Gall Line, Sfriso Adriano, Ceregato Alessandro, 2015. "Discovery of a forgotten historical algarium for the Venice Lagoon: the Vatova collection", 6th European Phycological Congress. London. *European Journal of Phycology*, vol. 50, pp. 124-125.
- Atlante della Laguna <http://www.atlantedellalaguna.it> (26 giugno 2017).
- Balletti Caterina, Ceregato Alessandro, Gottardi Caterina, Rizzi Francesca, Vianello Andrea, 2016. "3D Digitization and Web Publishing of an ISMAR Cartographic Heritage: Historical Maps of Venice Lagoon" *e-Perimetron*, vol. 11, n. 2, pp. 47-56.
- Begnoche Michelle, 2002. "Specimens at herbarium get digital makeover", *The University Record Online*, http://ur.umich.edu/0102/Nov11_02/21.shtml (26 giugno 2017).
- Bertolla Giorgio, Birello Giancarlo, Perin Anna, 2012. "Digibess, una biblioteca digitale open source", *Biblioteche Oggi*, vol.30, n.6, pp. 25-30.
- Biodiversity Information Standards – TDWG. <http://www.tdwg.org/>. (26 giugno 2017).
- Ceregato Alessandro, 2015. "Luigi Ferdinando Marsili: militare, diplomatico e scienziato al servizio dell'imperatore e di se stesso e il suo contributo all'approfondimento delle conoscenze dell'assetto fisico del territorio tra l'Adriatico e il basso Danubio", in Grigore Arbore Popescu and Caterina Dabalà (Editors), *The cartography of the Danube and the surrounding areas in and after the times of Vincenzo Maria Coronelli and Luigi Ferdinando Marsili*, Venice, CORILA, pp. 54-63.
- Ceregato Alessandro, Armeli Minicante Simona, Minuzzo Tiziano, Birello Giancarlo, Perin Anna, 2016. "Algarium Veneticum. Da una collezione storica alla creazione di un archivio digitale multitematico", *Conferenza GARR, Firenze*, <http://eventi.garr.it/it/documenti/conferenza-garr-2016/paper/30-conf2016-paper-ceregato>.
- Dublin Core Metadata Element Set, <http://dublincore.org/documents/dces/>, versione tradotta in italiano dall'Istituto Centrale per il Catalogo Unico: http://www.iccu.sbn.it/opencms/opencms/it/main/standard/metadati/pagina_116.html (26 giugno 2017).
- Darwin Core Terms <http://rs.tdwg.org/dwc/terms/> (26 giugno 2017).
- OpenDOAR The Directory of Open Access Repositories, Policies Tool <http://www.openoar.org/tools/en/policies.php> (26 giugno 2017).
- Portale cartografico BSA <http://cigno.ve.ismar.cnr.it/bsa/> (26 giugno 2017).
- Schmidt Larry, 2007. "Digitization of Herbarium Specimens, a Collaborative Project". *Conference Proceedings of the Association of College & Research Libraries*, pp.64-69.

7 GLOSSARIO

Algario

Raccolta di cartoncini (*exsiccata*) sui quali vengono posizionate e fatte essiccare le alghe.

Collezione (per repository)

Raccolta ordinata di oggetti della stessa specie, che abbiano valore o per loro pregio intrinseco o per loro interesse storico o artistico o scientifico o semplicemente per curiosità o piacere personale (*da Enciclopedia Treccani online*). Nell'ambito del repository, per "collezione" si intende un insieme organico e congruente di oggetti.

Creative Commons

Le licenze Creative Commons, a cui ci si può riferire anche come *le Creative Commons*, sono delle licenze di diritto d'autore redatte e messe a disposizione del pubblico a partire dal 16 dicembre 2002 da Creative Commons (CC), un'organizzazione non a scopo di lucro statunitense fondata nel 2001 da Lawrence Lessing, professore di diritto all'Havard University.

Queste licenze si ispirano al modello copyleft già diffuso negli anni precedenti in ambito informatico e possono essere applicate a tutti i tipi di opere dell'ingegno.

Le licenze CC, in sostanza, rappresentano una via di mezzo tra copyright completo (*full-copyright*) e pubblico dominio (*public domain*): da una parte la protezione totale realizzata dal modello *all rights reserved* ("tutti i diritti riservati") e dall'altra *no rights reserved* ("assenza totale di diritti").

La filosofia su cui si fonda lo strumento giuridico delle licenze CC si basa sul concetto *some rights reserved* ("alcuni diritti riservati"): in questo senso è l'autore di un'opera che decide quali diritti riservarsi e quali concedere liberamente (*da Wikipedia*).

Darwin Core (DWC)

Darwin Core (spesso abbreviato DwC) sono un'estensione dei metadati Dublin Core specifici per la biodiversità. Gli elementi descritti nello standard sono una parte di un più ampio set di vocabolari e specifiche tecniche sviluppate e curate dal Biodiversity Information Standards (TDWG) (già conosciuto come Gruppo di lavoro su database tassonomico (TDWG) (*tradotto in Italiano da Wikipedia*).

Dublin Core (DC)

I Dublin Core (dal nome della città americana nell'Ohaio) è un sistema di metadati costituito da un nucleo di elementi essenziali ai fini della descrizione di qualsiasi materiale digitale accessibile via rete informatica (*da Wikipedia*).

Fedora

Fedora è un sistema robusto, modulare, open source per gestire repository per la disseminazione di contenuti digitali. È particolarmente pensato per biblioteche e archivi digitali sia per l'accesso che per la preservazione. È anche utilizzato per accessi specifici a collezioni digitali storiche o culturali molto ampie e complesse e ai dati della ricerca. Fedora è installato in tutto il mondo da organizzazioni culturali e accademiche, università, enti di ricerca, biblioteche universitarie, biblioteche nazionali e enti governativi (*tradotto da <http://fedorarepository.org/>*).

Ingesting o/data ingestion

Ingesting o data ingestion è il processo di acquisizione e di importazione dei dati per l'utilizzo immediato o l'archiviazione. I dati possono essere trasferiti in tempo reale o trasferiti in batch ossia importati in blocchi a intervalli di tempo periodici.

Un processo efficace di ingestione di dati deve prevedere le priorità delle fonti, la convalida dei singoli file e l'inserimento nella destinazione corretta (*tradotto da <http://whatis.techtarget.com>*).

Islandora

Islandora è un framework open source progettato per aiutare le istituzioni e le organizzazioni e i loro utenti a gestire e a scoprire le risorse digitali utilizzando le migliori pratiche. Islandora è stato originariamente sviluppato dalla Biblioteca Robertson

dell'Università Prince Edward Island, ma è utilizzato e costantemente migliorato da una comunità internazionale sempre crescente (*tradotto da <https://islandora.ca/about>*).

Metadati

Un metadato (dal greco μετά “oltre, dopo, per mezzo” e da latino *datum* “informazione” - plurale: *data*), letteralmente “(dato) per mezzo di un (altro) dato”, è un’informazione che descrive un insieme di dati (*da Wikipedia*).

PID o Persistent identifier

I persistent identifier sono degli identificatori univoci di oggetti digitali che permettono un accesso durevole nel tempo alle risorse elettroniche, indipendentemente dalla loro collocazione.

Policy

Le policy di un repository sono delle linee di condotta dichiarate, in particolare per i repository Open Access è importante renderle pubbliche per il riutilizzo di dati e metadati così come per la sottomissione dei contenuti e la loro preservazione nel tempo.

Portale

Con particolare riferimento alla rete Internet, un sito che offre all’utente una serie di servizi e informazioni provenienti da più sorgenti, fornendo collegamenti ai siti che li offrono e collegamenti tematici per un vasto numero di argomenti. Un portale non differisce quindi sostanzialmente da un motore di ricerca del tipo directory, con il quale talvolta si identifica. Il vantaggio per l’utente consiste nel disporre di un unico sito sul quale trovare tutte le informazioni desiderate (*da Enciclopedia Treccani online*).

Repository

Un contenitore sicuro e affidabile di oggetti digitali (ad esempio libri, foto, audio, video, dati della ricerca, ecc.), corredati da metadati.

Tassonomia

È la scienza che nomina, descrive e classifica gli organismi. Il termine *taxon* (pl. *taxa*) è impiegato genericamente per designare un raggruppamento sistematico di qualsiasi rango (specie, genere, famiglia ecc.) (*da Enciclopedia Treccani online*).

Thumbnail

parola inglese traducibile in “miniatura”, in informatica è il termine che viene usato normalmente per definire l’anteprima di un’immagine più grande, che quindi viene presentata in formato ridotto (*adattato da Wikipedia*).